DOCUMENT RESUME

ED 235 608 EC 160 444

AUTHOR Reuter, Jeanette; And Others
TITLE Caregiver Reports on the Developmental Status of Handicapped Young Children: The Kent Infant Development Scale and the Minnesota Child Development Inventory.
INSTITUTION Kent State Univ., Ohio.
SPONS AGENCY Special Education Programs (ED/OSERS), Washington, DC.
PUB DATE Nov 82
GRANT G008001794
NOTE 54p.; Paper presented at the Annual Conference of the Association for the Severely Handicapped (9th, Denver, CO, November 4-6, 1982).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Behavior Rating Scales; Developmental Stages; *Evaluation Methods; Moderate Mental Retardation; Primary Education; *Severe Disabilities; *Test Reliability; Test Use; *Test Validity; Young Children
IDENTIFIERS *Kent Infant Development Scale; *Minnesota Child Development Inventory

ABSTRACT

This panel presentation presents results of an assessment study of the reliability, validity, and utility of caregivers' reports on: (1) the behavioral competencies of severely handicapped children, and (2) the adaptive and intellectual behaviors of moderately handicapped children. The Kent Infant Development (KID) Scale (used with severely and profoundly handicapped children) and the Minnesota Child Development Inventory (MCDI) (used with moderately handicapped children in a parallel matrix of testing) were studied. The KID Scale elicits caregiver responses on child competencies in five areas: cognitive, motor, language, self-help, and social. Data are used for computer-generated profiles, including developmental timetables that indicate which milestones have been acquired and which milestones should be acquired next. Analysis of KID's reliability indicates adequate interjudge and test-retest reliability. Studies on the test's validity established concurrent validity with the Bayley Scale of Infant Development and substantiated the validity of caregiver reports. The scale's utility is discussed, and its prescriptions for programming are emphasized. The MCDI uses the mother's observations to measure development in eight areas: general, gross motor, fine motor, expressive language, comprehension-conceptual, situation comprehension, self-help, and personal-social. In this study, the instrument was completed by both home and educational caregivers and results were compared with the Stanford Binet Mental Age measure for 93 moderately retarded primary school children. Results indicated that the General Developmental scale of the MCDI was the best measure of Developmental Age in terms of reliability and validity, had the highest interjudge correlation, and had the highest correlation with the Stanford Binet Mental Age. (CL)

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED235608

Caregiver Reports on the Developmental Status of Handicapped Young Children: The Kent Infant Development Scale and the Minnesota Child Development Inventory

A Symposium presented at the Ninth Annual TASH Conference

Denver, Colorado

November 5, 1982

by

Jeanette Reuter, Ph.D., Chmn.

Virginia Dunn, M.A.

Terry Stancin, M.A.

James Moe, Ph.D.

from

Kent State University and Kent Developmental Metrics, Inc.

P.O. Box 3178

126 West College Avenue

Kent, Ohio 44240-3178

(216)-678-3589

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jeanette Reuter

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EC160444

Symposium Outline

Introduction..... p. 1

Jeanette Reuter

The Reliability of the KID Scale..... p. 7

Virginia Dunn

The Validity of the KID Scale..... p. 16

Terry Stancin

The Utility of the KID Scale..... p. 28

Jeanette Reuter

The Minnesota Child Development Inventory..... p. 35

James Moe

Introduction

Jeanette Reuter, Kent State University

Caregivers' information about the developmental status of severely handicapped children has been the object of study of a two year research grant to the First Chance Project-Research at Kent State University by the Office of Special Education.[1] This panel presentation is the first complete report on the results of an assessment study whose goals were to demonstrate the reliability, validity, and utility of caregivers' reports on the behavioral competencies of severely handicapped children.

Research Design: First Year

During the first year, the Kent Infant Development (KID) Scale (Katoff, Reuter & Dunn, 1980) was successfully adapted to elicit reliable developmental information from the mothers, teachers, nurses, therapists, and child care workers of 121 severely handicapped children. To test the validity of that information, it was compared to the developmental information provided by the Bayley Scales of Infant Development (Bayley, 1969) on each child. Computer-based procedures for interpreting the KID Scales led to their application in the design of individual habilitation programs and for following the developmental progress of each research child.

The KID Scale contains 252 items in the form of phrases describing behaviors characteristic of an infant in its first year of life. On the basis of content, items are divided into five domains: cognitive, motor, language, self-help, and social. A caregiver marks on an answer sheet those behaviors she has seen her child perform. A computer program reads the responses from the optically scanned answer sheet, prints out the items in order of developmental age by domain, and compares the results for each domain and for the full scale with the results of the 500 healthy infants in the normative sample. The printout furnishes developmental ages

---

[1] This study was supported by Department of Education Research Grant DED-G008001794. Views expressed herein do not necessarily represent those of the Department of Education.

for each domain and for the full scale, a profile of strengths and weaknesses, as well as a developmental timetable showing what developmental milestones have already been acquired and indicating those to be acquired next. This timetable makes a direct bridge from developmental assessment to prescriptive programming and forms the basis for a caregiver/professional conference.

A comprehensive matrix of testing provided the data required to adopt an infant behavior inventory, the KID Scale, for use with severely handicapped young children and to establish psychometric standards for that adaptation. Pilot reliability and validity studies were first carried out utilizing a data pool of KID Scales which had accumulated from the evaluation of demonstration and outreach research conducted on severely handicapped children in the preceding six years. This work suggested the feasibility of demonstrating the reliability and validity of a caregiver report inventory on the behavioral repetoires of severely handicapped young children. For this purpose, new data were gathered from the caregivers of severely handicapped young children. Ann Copeland, Ph.D., in Massachusetts; Katherine Reuter, Ph.D. in California; Cindy Legin-Bucell, Ph.D. in Georgia, assisted Jeanette Reuter and Virginia Dunn in locating and testing 121 severely handicapped young children and two caregivers for each of them. The testing matrix included two KID Scales two weeks apart by each caregiver of each child with a Bayley Scale of Infant Development administered to each child during that two week interval. This arrangement of tests allowed for calculating test-retest reliability of the KID Scale over a two week interval, for assessing the interjudge reliability of two caregivers' KID Scales by domain scores and item by item in a percent agreement format, and for the calculation of the correlation of KID Scale and Bayley Scale DA's, a validity coefficient. After an interval of six months, each caregiver filled out a KID Scale on the severely handicapped child in her charge allowing for a short term follow-up. Table 1 summarizes the study samples drawn from the KID Scale testing matrix just described. It should be noted that the sample size of the studies varied due to independence considerations, and caregiver attrition.

Table 1

Study Samples for the Adaptation of the

KID Scale for Severely Handicapped Children

1. PILOT DATA

Resource: Clinical KID Scale data pool accumulated 1978-1981

2. RESEARCH DATA

Resource: Prospective research data accumulated 1980-1982

Individual Caregivers	Pairs of Caregivers
Test/Retest Correlation Sample (N = 121 children, 121 caregivers) A KIDS completed for each child by a caregiver on 2 occasions with each child and each caregiver used only once.	Interjudge/Correlation Sample (N = 112 children, 224 caregivers) A caregiver could report on as many as 3 children but each child and each caregiver pair is unique.
Item Age Norm Validity Sample (N = 121 children, 121 caregivers) Same as above except that it is a proper subset of the interjudge % agreement sample.	Interjudge/Percent Agreement Sample (N = 112 children, 224 caregivers) Same as above.
Concurrent Validity Sample (N = 106 children, 106 caregivers) Most reliable caregiver's second KID Scale and BSID scores.	

Research Design: Second Year

The feasibility of extending this model using caregiver information to design and evaluate habilitation programs for older, moderately retarded children was determined by studying a parallel matrix of testing with the Minnesota Child Development Inventories (MCDI) (Ireton & Thwing, 1974) completed by the mothers and teachers of moderately retarded children aged five to ten years. The reliability and validity coefficients of the developmental observations of caregivers proved to be substantial when tested against the Stanford Binet Intelligence Scale and the McCarthy Scales of Children's Abilities.

The MCDI is a standardized instrument for using the mother's observations to measure the development of her child. The Inventory is to be used with children from one to six years of age and contains 320 items, grouped into eight developmental scales: General Development, Gross Motor, Fine Motor, Expressive Language, Comprehension-Conceptual, Situation Comprehension, Self Help, and Personal-Social. Using the KID Scale computer scoring as a model, a similar MCDI computer scoring format was developed. This computer program reads the responses from the optically scanned sheet and prints out developmental ages (DA) for each scale yielding a profile of strengths and weaknesses as well as a developmental timetable similar to the one for the KID Scale.

During the second year of the research in order to test the value of using caregiver reports to assess moderately handicapped children in the primary grades in school, the caregiver report from the MCDI was correlated with a validity criterion, the Stanford Binet Test. The subject pool consisted of 93 children each with an MCDI report from a home caregiver and an educational caregiver. The interjudge reliability and the validity of the MCDI DA's were investigated.

Four advantages will be gained if caregivers' observations of developmentally significant behaviors meet psychometric criteria: no untestable children, cost efficiency, a rich record of functional behaviors to be used in programming,

and early caregiver involvement in habilitation designs. First, the fact that children are observed in their own environment by familiar caregivers over many hours of intimate caregiving provides insurance that they will be given every opportunity to demonstrate all of their competencies. This is in contrast to the limited test period in which an unknown examiner often in unfamiliar surroundings requires a child to perform nonfunctional sample behaviors on command. Second, cost efficiency results from substituting non-professional time for professional time in obtaining the developmental assessment data on a child.

The third advantage derives from the wealth of specific behavioral information which can be obtained from the use of caregiver inventories. The two inventories mentioned above yield not just developmental ages--the usual product of infant tests--but a profile of strengths and weaknesses and a developmental timetable showing which developmental milestones will be acquired next. This timetable makes a direct bridge from developmental assessment to prescriptive programming and forms the basis for a caregiver/professional conference.

The fourth advantage grows out of the early involvement of those most responsible for the successful follow through, the caregivers, in the habilitation design. One can hope that this early direct involvement in developmental observations will enhance the prospects for a more felicitous implementation of developmental programming on the part of caregivers.

However, caregiver information must be reliable and valid in order to realize any true advantage from its use. The results of the present research reveal substantial reliability and validity, and suggest by what means and under what circumstances caregiver information can be useful.

References

Bayley, Nancy. Bayley Scales of Infant Development. New York: The Psychological Corporation, 1969.

Ireton, H. & Thwing, E. J. The Minnesota Child Development Inventory. Minneapolis, MI: Behavior Science Systems, Inc., 1974.

Katoff, L., Reuter, J. & Dunn, V. The Kent Infant Development Scale Manual. Kent, Ohio: Kent Developmental Metrics, 1980.

Reliability of the Kent Infant Development Scale

Virginia Dunn, Kent State University

Introduction

Developmental information essential to the planning of care and instruction of severely and profoundly handicapped young children must reflect their demonstrated capacities. However, assessing their capabilities accurately requires observation and an objective tool to record and quantify those observations. The tool under study here is the Kent Infant Development Scale. In order to establish the reliability of this scale, it has been necessary to examine the consistency with which observers report their information about a child, since these reports are used in establishing what a child can and cannot do. In the past, most data concerning the limitations or competencies of such handicapped children have been produced for the most part from the observations of professionals. Now, however, a valuable source of information about any child--the primary caregiver--is being tapped. This new source is of particular advantage in the case of children with limited behavioral competencies. Obviously, a direct caregiver has more opportunities to observe the full range of behaviors existing in the child's repertoire than does the professional, who is able only to spend an hour or so with the child in a novel situation. Further, an inventory of behaviors completed by the caregiver is less expensive than a professionally administered test. At the same time filling out the inventory increases caregiver involvement in all phases of the child's treatment. Therefore, our research over the last five years has had the objective of structuring caregiver reporting in such a way as to yield in an accessible format consistent and reliable information about severely and profoundly handicapped young children.

The present study makes use of information gathered from two sources: the pilot and the research data. The pilot data was compiled from KID Scales gathered over several years from caregivers responsible for residents of the Hattie Larlham Foundation, a residential treatment center for severely and profoundly handicapped children in Mantua, Ohio. The research data consisted of KID Scales from four different places across the country--California, Florida, Georgia, and Ohio--collected on 121 children. In each geographical area, a set of two caregivers completed a KID Scale on a child, followed by another testing in approximately two weeks. Then six months later at least one of the two original caregivers completed another KID Scale on the child.

This portion of the symposium reviews the consistency of caregiver reports as structured by the KID Scale. The first question under study was the degree of consistency with which any one caregiver can be expected to report on the behavior of a child under her care. The object here was to discover if caregivers would report much the same information on two different but closely spaced testing occasions. This is to say, would the data yield test-retest reliability. The next task was to discover the extent to which two caregivers, when observing the same child, will agree on their descriptions. This, in turn, was undertaken to establish the interjudge reliability of the scale.

Test-Retest Reliability

The purpose of establishing test-retest reliability is to determine if caregivers reporting on a child on two occasions separated by an interval of two weeks should be expected to produce the same developmental age on the second KID Scale as on the earlier completion. The test data were transformed to developmental ages and the correlations of the domain and full scale developmental ages from the KID Scales completed two weeks apart by each caregiver were computed with the results presented in Table 2. The caregivers were classified into three

Table 2

KID Scale Test-Retest Reliability Coefficients

by Caregiver Type

(N = 121 caregivers)

Caregiver Type	Cognitive	Motor	Language	Self-Help	Social	Full Scale
Parents (N=45)	.98	.99	.94	.97	.96	.98
Non Professionals (N=36)	.96	.98	.95	.92	.96	.99
Professionals (N=40)	.95	.99	.96	.97	.97	.99
Total	.97	.99	.95	.96	.96	.99

Table 3

KID Scale Test-Retest Percent Agreement Means

(N = 112 caregivers)

Cognitive	Motor	Language	Self-Help	Social
91%	93%	90%	92%	89%

groups: 45 parents, 36 nonprofessionals, and 40 professional caregivers. The resultant correlations for each group are uniformly greater than .92, both for domain scores or full scale scores, as well as for the results obtained from parents, nonprofessionals, or professionals. These correlations indicate that caregivers of any type who recorded low developmental ages during the first testing situation do the same in the second round. Similarly, those who reported high developmental ages on the first test did so on the second, also. Although the mean DA's are not reported here, no significant differences in DA's were present between the two testings. Thus, these results yield the assurance that if a teacher, an aide, or a parent observes his child on two separate occasions as structured by the KID Scale, the chances are excellent that the child would receive the same DA on both occasions.

Developmental age, of course, represents a summary statement about the developmental status of a child but is limited to providing a gross classification of developmental status. Further insight is provided by looking at how closely caregivers agreed with themselves on individual items over time. To accomplish this, caregiver responses to each item on both testing sessions were compared, resulting in the calculation of the mean percent agreement for each item across all caregivers, and finally an average percent agreement for each domain. These findings, found in Table 3, indicate that caregivers will agree with themselves on an item-by-item basis when they report their observations on the developmentally significant behaviors on two separate occasions. These two studies give us confidence that caregivers will report consistently on both the individual items and the global components of the KID Scale. Therefore, the reliability testing of this data indicates that the structuring of behavioral information in this way can elicit reliable responding.

A third test-retest study was conducted by Nancy Hoag of the Kent State University Psychology Department. In her study, she looked at 71 caregivers who had demonstrated reliable reporting. Six months after the initial pair of KID Scales, these caregivers completed KID Scales again on the same children. See Table 4. There was little difference between the two testing situations in the ordinal ranking of the scores, with the exception of a significant increase in the developmental status over the six month retest period. Again, the correlations are consistently above .90 while the mean developmental ages of the subsample of severely handicapped children were six months.

Interjudge Reliability Studies

In order to examine the extent to which two different caregivers working with the same child provide consistent information, the first interjudge study was carried out on the pilot data. The question that needed to be examined was the possibility that caregiver information, might, in part, be dependent on the educational background of the individual completing the scale. Table 5 is designed to address this question. The data presented here had been collected over the preceding six years from observations of residents at the Hattie Larlham Foundation in Ohio. All repeated KID Scales on any one child, collected within two months of one another, were used, and the results of these first and second KID Scales were correlated. Note in Table 5 that the interjudge reliability coefficients are lower than the test-retest reliability coefficients presented earlier for the KID Scale. This is the normal expectation when comparing these two types of reliability coefficients. The correlation on a full scale basis is .96 when two professionals were the observers. The correlation of full scales when direct caregivers and a professional form a pair was significantly lower at .86. When the pairs were composed of two direct caregivers, with no professional involved, the correlation of full scales became significantly lower at .82. Thus, when one professional is involved in the interjudge pair, the full scale correlations

Table 4

The Correlations, Mean Changes and t's of KID Scale DA's at Six Month Follow-Up

(N = 71 caregivers and children)

	Initial $\bar{X}$ Age in mo.	Follow-Up $\bar{X}$ Age in mo.	t	Correlation
Cognitive	6.3	6.6	1.5	.95
Motor	6.0	6.2	1.74*	.97
Language	6.2	6.7	2.73*	.90
Self-Help	7.3	7.4	0.95	.95
Social	6.2	6.6	2.45*	.94
Full Scale	6.3	6.7	3.49**	.97

*p < .05
**p < .001

Table 5

Interjudge Reliability Coefficients

by Caregiver Type

(Pilot Data)

Pair Type	Cognitive	Motor	Language	Self-Help	Social	Full Scale
Direct Caregivers (N=22 pairs)	.81	.76	.72**	.76	.81*	.82**
Direct Caregivers and Prof. (n=21 pairs)	.89	.91	.69**	.92	.67*	.86**
Professionals (N=21 pairs)	.93	.98	.83**	.92	.95*	.96**

*p < .01 Chi square tests for the significance of the differences between
**p < .001 independent correlations

are at least .86; but even when the pair consist of two direct caregivers, the correlation is still above .82. Since data were not available in our research sample to classify the caregivers according to their professional status, the only direct comparison of interjudge reliability by caregiver type comes from the pilot data.

The research study replicated and expanded the pilot study. Two separate KID Scale testings given two weeks apart were evaluated. The results (Test 1 and Test 2) calculated for the interjudge reliablilty on the 121 caregiver pairs for both testings appear in Table 6. The full scale interjudge reliability was similar to that calculated for the same study done with the pilot data. The reliabilities of motor, self help, and cognitive domain DA's were also at least .82 or above. On the social scale the correlations were .73 to .76, not quite as high; and on language they were just .70. Of course, the greater the number of items, the better chance there is for high reliability so the full scale reliabilities are always higher than subscale reliabilities.

Again, a mean percent agreement for each domain was calculated on an item by item basis in order to examine the degree to which members of caregiver pairs agreed with each other on individual items. This is a rigorous method of calculating interjudge reliability, with an 85% agreement between judges usually representing a good level of percent agreement for behavioral observations. As presented in Table 7, the percent agreement on the pilot sample ranged from a low of 85% on the Self Help Scale to 92% on the Motor Scale. The research samples' results are somewhat lower than the pilot data. The lowest percent agreement was 71% on the Social Scale and 72% on the Language Scale. The average percent agreement across all samples and on all studies reached 82%.

In summary, the data indicate that caregiver information will yield adequate interjudge and test-retest reliability. These reliability findings can be added to the growing body of evidence that demonstrates that the

developmental information from the people who work with and care for severely handicapped children can be trusted. Caregivers may be seen as reliable developmental observers of their charges when their observations are structured by the Kent Infant Development Scale.

Table 6

KID Scale Interjudge Reliability Coefficients

(N = 121 caregiver pairs)

	Cognitive	Motor	Language	Self-Help	Social	Full Scale
Test 1	.84	.95	.69	.91	.76	.89
Test 2	.82	.94	.71	.88	.73	.87

Table 7

Interjudge Percent Agreement Means for KID Scale Items

	Cognitive	Motor	Language	Self-Help	Social	Full Scale
Sample Pilot (N=52 pairs)	89%	92%	86%	85%	86%	87%
Test 1 Research (N=112 pairs)	82%	82%	85%	85%	80%	82%
Test 2 Research (N=112 pairs)	75%	82%	72%	79%	71%	76%

Validity of the Kent Infant Developmental Scale

Terry Stancin, Kent State University

The next segment of this symposium explains the process by which the validity of the KID Scale was established. In psychometric research, establishing validity is necessary to determine if the given test is indeed measuring that for which it was designed and then to discover how well it performs that function. One method for determining the validity of a developmental assessment devise is to compare the results from the newer test with those obtained on another, well-established test. If the two sets of scores are consistent for a large sample of children, if they correlate, then one may assume that the tests are measuring the same phenomena and concurrent validity has been established. In part, concurrent validity establishes the psychometric properties and appropriate uses of a test.

A further issue of concern in establishing the psychometric integrity of the KID Scale involves the validity of caregiver-based testing in general. Some professionals have questioned the desirability, credibility, and accuracy of caregiver reports, particularly those made by parents, who they assume to be positively biased and lacking the objectivity necessary for accurate observation and valid reporting. Of note, however, are the findings from most of the studies making use of caregiver reports suggesting that they can be seen as reliable observers of their children's contemporaneous behaviors and developmental functioning, for example, Gradel, Thompson and Sheehan (1981) and Kaplan and Alatishe (1976). In keeping with this, it has been found that maternal reports are stable over time and also highly correlated with the observations of other caregivers and professionals. However, some studies have demonstrated that caregivers, particularly mothers, predict that their handicapped child can perform a greater number of behaviors on a psychological test than, in fact, the child subsequently

performs for a professional examiner. Therefore, maternal reports yield higher developmental estimates than do professionally administered tests. The researchers in most of these studies, however, did not use a standardized, psychometrically sound inventory. Nonetheless, these researchers concluded that even while mothers seem to "overestimate" their child's developmental status, they can be used as reliable and valuable sources for developmental information (Stancin, 1981).

Concurrent Validity with Bayley Scales

The following two studies were designed to examine these validity issues with respect to the KID Scale. In the first study, scores from the KID Scales completed by caregivers of handicapped children were compared to test scores from professionally administered Bayley Scales of Infant Development. The Bayley is the most frequently used standardized developmental test for assessing severely handicapped young children functioning at an infant developmental level. The purpose of this study was to examine concurrent validity. In the second study, KID Scale reports from mothers were compared to those from teachers as well as to Bayley results. This study allowed for the examination of the differential reporting of caregivers on a structured inventory, the KID Scale.

Subjects for both studies were selected from the population of severely handicapped young children previously described. As stated before, two caregivers provided developmental information for each child by completing two KID Scales on a child within about a two-week interval. During the same interval, the Bayley Scales were professionally administered to each child. Of the two completed KID Scales, the second was selected to compare with the Bayley.

The sample used for Study 1 was constructed to reduce the effects of unreliable reporting on validity and to derive measures of validity based on independent caregiver-child pairs. For this reason, the most reliable caregiver of the two available for each child was selected for this study. One hundred and six children from the data pool and their most reliable caregivers formed

the subject-caregiver pairs. Of the children, 61 were male and 45 female between the ages of 18 months and nine years, for which the mean age approximated five years. The caregivers were comprised of 35 mothers, 30 teachers, 25 child care aides, a grandmother, and a ward nurse.

Table 8 lists the concurrent validity coefficients between the KID Scale domain scores and the Bayley Mental and Motor Scale scores. Seventy-five percent of these coefficients are greater than .80, which makes them significant and acceptable in psychometric terms. The KID Scale full scale scores correlated with the Bayley Mental and Motor scales at .85.

Table 9 lists the differences between the mean developmental ages derived from the KID Scale and Bayley Scales. Although scores are highly correlated, the KID Scale developmental age estimates are about one month greater than the developmental age estimates derived from the Bayley Scales. However, this month difference is not clinically significant and is probably due to the norming procedures for the two tests.

There are two differences in the construction of DA norms between the KID Scale and the BSID. First, chronological age designations were calculated differently for the different norming samples. KID Scale ages were based on the infant being in its nth month; i.e., a three month label included infants between the ages of two months, 1 day, to exactly three months. Thus an age of three months on the KID Scale has a midpoint age of 2 1/2 months. However, the BSID age norms were based on a sample of infants who were tested at the given age within a four day limit on either side yielding a midpoint age of three months in the above example. As a consequence, KID Scale norms result in age labels that are about one half month higher than the BSID labels.

A second norming construction difference is the criteria for determining item age designations. The KID Scale item age norms on which DA's are based were the age at which 65% of the children of an age passed a given item, whereas

Table 8

Raw Score and DA Validity Coefficients of the KID Scale and the BSID Derived from Mother and Teacher Reports[a,b]

KID Scale Domains	BSID Scales			
	Mental		Motor	
	Raw	DA	Raw	DA
Cognitive	.878	.844	.850	.789
Motor	.843	.803	.957	.912
Language	.737	.707	.634	.597
Self Help	.854	.804	.875	.794
Social	.801	.764	.720	.693
Full Scale	.885	.851	.897	.857

[a] N = 106.

[b] All $p < .001$.

Table 9

Means, S.D.'s and differences Between DAs from the KID Scale and the BSID in Months[a]

KID Scale Domains			Difference from BSID Scales			
			Mental[b]		Motor[c]	
	Mean	S.D.	D[d]	D[e]	D[d]	D[e]
Cognitive	6.3	4.5	+1.1	+0.1	+0.6	-0.4
Motor	6.1	4.0	+0.9	-0.1	+0.4	-0.6
Language	6.3	3.3	+1.1	+0.1	+0.6	-0.4
Self Help	7.4	4.0	+2.2	+1.2	+1.7	+0.7
Social	6.3	3.5	+1.1	+0.1	+0.6	-0.4
Full Scale	6.4	3.6	+1.2	+0.2	+0.7	-0.3

[a] N = 106

[b] BSID Mental Scale DA: Mean = 5.2, S.D. = 4.3.

[c] BSID Motor Scale DA: Mean = 5.7, S.D. = 5.5.

[d] Differences between KIDS DAs and BSID DAs.

[e] DA differences corrected for CA and passing criterion differences (1 month).

a 50% passing criterion was used with the BSID. The 65% passing criterion of the KID Scale requires, on the average, 10 items less than the 50% criterion to attain a specific DA. This is about 50% of the nineteen items required, on the average, to move from one DA to a month higher. Thus, the combined effect of these two factors is that KID Scale DA's are one month higher than Bayley DA's due to these normative differences. This constant displacement does not effect the validity coefficients of the KID Scale and the BSID but it is necessary to reduce KID Scale DA's one month when comparing them directly to BSID DA's.

Thus the results, as summarized in Tables 8 and 9, suggest that as a group, caregivers report information on the KID Scale that is consistent with the information obtained by professionals when using the Bayley Scales on the same handicapped children.

Concurrent Validity of Caregivers

The second study examined the differential validity of KID Scale reporting of mothers and teachers. In Study 2 all available pairs of mothers and teachers or teachers' aides were drawn from the data pool. This resulted in 57 independent mother/teacher pairs of caregivers reporting KID Scale information on a child. The interjudge reliability coefficient between these mothers and teachers were all highly significant and ranged from .68 to .96 across domains.

Table 10 contains the validity coefficients for the mothers and teachers. These are the correlations between KID Scale scores and the Bayley scores for the mothers and the teachers, separately. Both sets of coefficients are high with teachers consistently somewhat higher, indicating that scores on KID Scales completed by mothers and teachers are highly related to Bayley results as well as to each other.

In Table 11, the mean developmental ages (DA's) for the two caregiver groups derived from the KID Scale are compared. In every domain the mothers' KID Scales

Table 10

DA Validity Coefficients

for the Mothers and Teachers[a,b]

	KID Scale Domains	BSID Scales Mental	BSID Scales Motor
		DA	DA
Mothers	Cognitive	.816	.781
	Motor	.751	.901
	Language	.624	.502
	Self Help	.764	.800
	Social	.757	.702
	Full Scale	.814	.851
Teachers	Cognitive	.872	.768
	Motor	.820	.932
	Language	.788	.586
	Self Help	.818	.842
	Social	.819	.698
	Full Scale	.890	.871

[a] $N = 57$.

[b] All $p < .001$.

Table 11

A Comparison of KID Scale DA Estimates in Months from Mothers' and Teachers' Reports[a]

KID Scale Domains	Mothers		Teachers			
	Mean	S.D.	Mean	S.D.	D[b]	t
Cognitive	7.6	4.5	6.3	4.0	1.3	4.4***
Motor	7.2	4.3	6.5	4.2	0.7	4.1***
Language	7.7	3.0	6.3	3.3	1.4	4.1***
Self Help	8.6	3.7	8.0	3.8	0.6	2.7**
Social	7.5	3.3	6.2	3.3	1.3	4.3***
Full Scale	7.6	3.5	6.6	3.5	1.0	4.9***

[a] $N = 57$, $df = 56$.

[b] Difference in DAs in months

**$p < .01$.

***$p < .001$.

yielded significantly higher DA estimates than did the teachers' KID Scales. Thus, while there were no discernable differences in the validity coefficients for the two caregiver groups, there were differences in mean estimates of developmental age. These results replicate findings of previous research that reported higher DA estimates by mothers than teachers. The high inter-judge correlations between mothers' and teachers' KID Scales, and their similar high concurrent validity coefficients are an indication that both sources of information are reliably reporting on similar behavioral observations. However, mothers are reporting that they observe more behaviors than teachers do.

Discussion

In conclusion, note that every concurrent validity coefficient of the KID Scale with the Bayley is significant and high. The small differences in DA's obtained from the Bayley and the KID Scale depend primarily on the norming procedures used for both tests and somewhat on the caregiver role of the reporter. These differences, approximately one month, derive from the age norming displacement between the two tests. The results obtained from this research support the validity of caregiver reports of developmental information for severely handicapped young children. Particularly, the results substantiate the KID Scale's diagnostic utility and interchangeability with respect to the Bayley. Therefore, with severely handicapped young children, a clinician can obtain developmental information from the KID Scale as a substitute for the Bayley Scales.

The diagnostic equivalency between the two tests carries with it significant implications. The lower administrative costs of caregiver completed instruments permit more frequent assessments, thus facilitating treatment planning and evaluations. In view of this more efficient tool, the psychologist's contribution to the assessment can be concentrated on the intervention and interpretation

phases rather than on the test administration and scoring process. The KID Scale contains functional, observable behaviors as items. For this reason, it has ecological validity, yielding prescriptive utility. The functional items describe competencies that the children need to learn. Conversely, it is of little adaptive value to teach a child the Bayley items.

Mothers endorsed more KID Scale items on their children, and, therefore, their reports yielded slightly higher estimates of developmental status than did those of teachers. There are several possible explanations for these discrepancies. Mothers have more experience with their children over longer periods of time than do teachers. This may give them more opportunity to observe developing behaviors which they can then endorse on the KID Scale. Therefore, summative competency judgments that mothers make on the KID Scale are based on more extensive behavioral sampling than the judgments made by teachers. The other explanation for the discrepancies posited by earlier researchers is that the mothers lack objectivity and, therefore, overestimate their children's abilities. The concrete, behavioral nature of the KID Scale argues against the overestimation hypothesis for this study because caregivers do not make predictions about how a child will respond to test items. Rather, they simply state whether a specific behavior has ever been observed to be in the child's repetoire. In addition, the KID Scale items are presented to the caregiver in a random order with respect to item age norms and domain content, making consistent overestimation difficult. It is important to remember that KID Scale items are normed to mothers' reports, thereby statistically compensating for any such bias should it occur.

For these reasons, whether or not mothers tend to overestimate their own children's behavioral competencies has now become a moot point. More to the point are our research findings that indicate that while they may report additional behaviors, little evidence exists that mothers are misrepresenting their perceptions

of their child's behaviors. In fact, mothers' perceptions are very similar to those of teachers and other caregivers. We know that successful early intervention programs must involve caregivers, both mothers and teachers, as much as possible. An effective way to ensure that caregiver participation is early and strong is to involve them in the initial and subsequent assessment activities. The KID Scale provides such an opportunity (Stancin, Reuter, Dunn, & Bickett, in press).

References

Gradel, K., Thomson, M. S. & Sheehan, R. Parental and professional agreement in early childhood assessment. Topics in Early Childhood Special Education, 1981, 1:2, 31-40.

Kaplan, H. E. & Alatishe, M. Comparison of ratings of mothers and teachers on preschool children using the Vineland Social Maturity Scale. Psychology in the Schools, 1976, 13, 27-28.

Stancin, T. The concurrent validity of the Kent Infant Development Scale. Master's Thesis, Kent State University, Kent, Ohio, 1981.

Stancin, T., Reuter, J., Dunn, V. & Bickett, L. The validity of caregiver information on the developmental status of severely and profoundly brain-damaged young children. American Journal of Mental Deficiency, in press.

The Utility of the Kent Infant Development Scale

Jeanette Reuter, Kent State University

In addition to the reliable and valid DA's and detailed descriptions by caregivers of the behaviors of infants and handicapped children available from the KID Scale, a printed report is provided which is the product of computer scoring and contains developmental information useful for designing habilitation programs for severely handicapped children. This computer printout contains a list of age-ordered items for each behavioral domain: cognitive, motor, language, self help, and social. The age ordering of the items is based on the age at which the healthy infants in the normative sample acquired the item behaviors, according to their mothers' reports. A section of the printout for the cognitive domain is reproduced in Table 12.

Those items which occur first in any domain on the printout have the lowest average age of acquisition and those which occur last in any domain were acquired latest. Thus, those items with the lowest age means are the easiest to acquire: those with the higher age means are harder. The printout for a healthy infant will have all A's or passes in the Checked column up to a certain point and then D's will appear mixed with the A's. Then a string of D's will appear continuing to the end of the domain. The area where D's and A's are mixed will occur approximately at the place where the item age means correspond to the baby's chronological age. The items in this area describe behaviors which the healthy infant will develop in the next few weeks, the area of emergent behaviors. Thus, the four items which are reported as the first four D's in a row on each of the five domains are 20 behaviors on which the infant will soon be "working" developmentally. Mothers can be alerted to these coming developments so that they can help their babies acquire them and so that they can reinforce the first approximations to these behaviors as they occur naturally.

Table 12

A Section of the KID Scale Cognitive Domain Printout for a Healthy Six Month-Old Baby

Item	Checked	Mean	Description
236	A	4.8	Tries to touch moving objects
201	A	4.9	Reaches for toys slightly out of reach
129	D	5.0	Reaches for everything in sight
27	A	5.2	Picks up objects and looks at them
106	A	5.4	Moves to get an object out of reach
63	A	6.3	Drops and picks up toys
245	D	6.5	Tries to catch moving objects
154	A	6.6	Drops toys and watches them fall
188	A	6.7	Plays with two toys at the same time
18	D	6.8	Overcomes obstacles to reach things
187	A	6.9	Smiles at the sight of a favorite toy
203	D	7.1	Smiles at the sight of a new toy
202	D	7.5	Drops one of two toys held to pick up a third
226	D	7.7	Looks for fallen objects by bending over
127	D	7.8	Finds half hidden objects
215	D	8.1	Squeezes dolls or toys to make them squeak

Note. Item = the number of the item as it appeared in the test booklet; checked = the mother's report on her child; mean = the item age mean; description = the item as it is written in the test booklet; A = yes; D = no, cannot do it yet.

The KID Scale printout can be used to build individual education and habilitation plans. A simple way to do this is to draw lines above and below the first four D's in a row on each domain and the resulting twenty behaviors can act as criterion behaviors for short term goals on an individual habilitation plan. Table 12 shows the correct position for these lines. As can be seen, this six month-old baby will be busily working on establishing object permanence and its mother will want to play hide and seek games with small manipulable objects with her baby this month.

From this it can be seen that the prescriptive use of the KID Scale printout focuses the mother's expectations so as to have the highest probability of successful reinforcement of her baby's developmental progress and, in turn, to have the highest probability of reinforcing her own motivations to elicit developmental progress.

What about the handicapped? Can the KID Scale printout be used prescriptively for severely handicapped young children as well as for normal infants? Is the item order established from the normative sequence of development for healthy babies the same as for severely handicapped young children? The following study by Virginia Dunn and me was designed to answer these questions empirically.

We decided to test whether the ordering of the domain behavior item inventories, obtained from KID Scale tests on the handicapped, ranked from the highest frequencies to the lowest frequencies of passing can be compared with the rank ordering of the behavior item inventories of normal infants by chronological ages. To do this the KID Scale items from the sample of severely handicapped children were rank ordered by domain with those items which the most severely handicapped children passed at the beginning of the domain list to those items which the fewest severely handicapped children passed at the end of the domain list. Unlike healthy infants, the chronological ages of severely handicapped children

do not approximate their development ages. Therefore, those items which only a few severely handicapped children passed were judged to be the hardest items and they should have the same rank order as the items that only the oldest babies passed. Those items which most of the severely handicapped children passed were the easiest items and should be the same items that the youngest babies in our normative sample passed. The item ranks established by the healthy babies' age of passing were compared with the item ranks established by the percent of severely handicapped children's passing by correlating their rank orders. These correlations can be found in Table 13. The high correlations indicate that the developmental item ordering established on healthy infants can be used for programming for severely handicapped young children without fear that the first four D's in a row will be inappropriate as a focus of habilitation. This is perhaps the most welcome result of our work. It makes it possible to recommend that the Kent Infant Development Scale can be used prescriptively for severely handicapped young children.

Thus, a start has been made toward establishing prescriptive use for the KID Scale which can be statistically supportive of clinical programming and caregiver reports for severely handicapped children. The behavioral content of the KID Scale items makes it useful then for describing a severely handicapped child's behavioral repertoire as it is at present and allows us to predict with some degree of accuracy what developmentally significant behaviors the severely handicapped child will be acquiring next. It gives us a list of criterion behaviors toward which Individual Habilitation Plans should be directed. As Stephen Porges suggests:

> Operant psychology, like developmental psychology, is dependent upon the observation, description, and measurement of changes in behavior over time. In order to select the behavioral sequence to be shaped, the operant psychologist is left with two options: (1) to systematically observe behavior and to decompose what appears to be normal (representative of the population) behavior into a series of

> elements which when combined through shaping procedures should result in the target behavior, or (2) to systematically observe the developmental sequence of a behavior and to shape behavior in accordance with this developmental framework. (1980, p. 758)

The KID Scale prescriptions are based on this second option.

Table 13

Correlations between percent endorsement of KID Scale items and the item age norm by domain order

	Cognitive	Motor	Language	Self-Help	Social
Healthy Infants	.99	.99	.99	.99	.99
Severely Handicapped Children	.91	.98	.91	.75	.89

References

Porges, S. Developmental designs for infancy research. In J. Osofsky, Ed., Handbook of infant development. New York: Wiley, 1980.

The Minnesota Child Development Inventory

James Moe, Kent State University

Introduction

The validity and utility of the caregiver completed Kent Infant Development Scale for assessing the developmental status of severely and profoundly handicapped children was demonstrated in the first year of this research described previously. The research in the second year explored the advantages and disadvantages associated with caregiver reports of adaptive and intellectual behaviors describing the developmental status of young, moderately retarded children in the primary grades of public special education classes. The caregiver report used was the Minnesota Child Developmental Inventory (Ireton and Thwing, 1974), and it was compared to the Stanford Binet (Terman and Merrill, 1973).

The Minnesota Child Developmental Inventory (MCDI) was chosen as the caregiver-completed instrument because the range of behaviors it covers is developmentally appropriate for describing moderately retarded young children. The MCDI consists of 320 statements which describe the behaviors of children from one to six and one half years of age. This range of behaviors corresponds to the developmental levels of five to ten year-old moderately retarded children. Caregivers record whether or not children display the behaviors described in the items on a yes/no format. The 320 items are divided, on a face content basis, into eight developmental domains: general development (GD), gross motor (GM), fine motor (FM), expressive language (EL), comprehension-conceptual (CC), self help (SH), and personal-social (PS). Developmental age levels are obtained for each of these domains.

The Stanford Binet was chosen as the best comparison instrument because of the wide age and ability range of its applicability and the solid base of psychometric research on its properties. The Wechsler Primary and Preschool Scale of

Intelligence (Wechsler, 1967) is appropriate for children between four and six and one half years of age and the Wechsler Intelligence Scale for Children--Revised (Wechsler, 1974) is appropriate for children between six and almost seventeen years of age. Therefore, the age ranges for neither of these two tests were developmentally or chronologically appropriate for our five to ten year-old moderately retarded sample of children. The McCarthy Scales of Children's Abilities (McCarthy, 1972) was not used because its appropriateness for assessing young, moderately retarded children has not been determined. Since the Stanford Binet (SB) has been widely criticized for its emphasis on verbal development, those items which require a verbal answer were separated from those which do not, and a separate mental age (MA) was calculated from each scale as well as the MA from the entire SB.

The primary purpose of this study was to determine the appropriateness of interchanging the MCDI for the SB for assessing the developmental status of moderately retarded young children. Interjudge reliability estimates were obtained on mother-completed and teacher-completed MCDI's. Concurrent validity estimates were obtained by correlating MCDI developmental age (DA) estimates with SB MA's and by comparing mean DA's with mean MA's. In addition to these primary concerns, two other issues were explored. The properties of the Language and Nonlanguage scales (developed for this study) of the SB were evaluated. Also, since the total sample had a preponderance of Down's Syndrome children, two subgroups were formed (comprised of Down's and non-Down's moderately retarded children) and the developmental age levels of these subsamples were compared.

Method

Four field consultants and home office staff located the children for this study. The consultants were: Fran Archer, Florida; Cindy Legin-Bucell, Georgia; Anne Copeland, Massachusetts; and Phil Piro, Ohio.

Subjects

Data was collected on 100 children. Seven children were eliminated from statistical analyses because of incomplete data or extreme deviation from the requirements for participation which defined the sample (IQ between 35 and 51, 5 to 10 years of age). The final sample consisted of 93 moderately retarded children, whose mean age was 98 months. The mean mental age was 43 months and the mean IQ was 43. Fifty-nine of the children were male and 34 were female. The children were attending a public school program or its equivalent while living at home or in a group home. Permission for participation in this study was obtained for each child from both the public school system and the child's legal guardian.

Ninety-three home caregivers completed MCDI's on the children. Seventy-six of the home caregivers were mothers, five were foster mothers, four were fathers, and eight were other home caregivers. The home caregivers had achieved an average of 12.4 years of education. The mean length of time caring for the children was 92 months.

Ninety-three educational caregivers completed MCDI's on their students. Sixty-four of the educational caregivers were teachers, 22 were teacher's aides, and seven were other educational caregivers. These caregivers averaged 15.5 years of education and the mean length of time caring for the children was 11 months.

Procedure

One of the four consultants from the research staff administered a Stanford Binet to each of the children of the experimental sample. Within two weeks of this test administration, two MCDI's were collected, one from each child's home caregiver and one from each child's educational caregiver. Thus each child had a unique pair of caregivers completing MCDI's. A psychological test report was

written based on the multi-source data and the reports were made available to the teachers and, through the teachers, to the parents.

Results

Interjudge Reliability

Reliability estimates were obtained by comparing MCDI scores on the same children from two different sources. Raw scores and developmental ages from parent and teacher reports for each of the developmental scales plus the Full Scale were correlated. Raw scores were simply the total number of endorsements within each scale. All correlations were Pearson product moment r's. Parent derived and teacher derived developmental age estimates for each developmental scale were compared with dependent t tests.

Correlations between parent derived and teacher derived raw scores for each developmental scale were highly significant (all p values $< .001$). The correlations ranged from .8785 for the General Development scale to .6271 for the Personal-Social scale (see Table 14). Developmental age estimates from parents and teachers were also highly correlated (all p's $< .001$). Again, the General Development scale showed the highest correspondence between parent and teacher reports with $r = .8657$ and the correlation for the Personal-Social scale was the lowest with $r = .5704$ (see Table 14).

Comparisons of mean developmental ages for each scale from parent and teacher reports show that parent estimates were typically higher than teacher estimates. Parent derived DA's were significantly higher than teacher derived DA's for all scales except for Gross Motor and Fine Motor (see Table 15).

Validity

Developmental age estimates from parent and teacher completed MCDI's were correlated with the criterion variable, Stanford Binet Mental Age (MA). In addition, to check the assumption that the MCDI was sensitive to developmental progression related to age for moderately retarded children, MCDI DA's were

Table 14

Correlations between MCDI Scores from Parent and Teacher Reports

Domains	Raw Scores	Developmental Ages
General Development	.8785	.8657
Gross Motor	.8021	.6252
Fine Motor	.6825	.6972
Expressive Language	.8142	.7603
Comprehension-Conceptual	.8430	.8116
Situation Comprehension	.6645	.6155
Self Help	.6754	.6687
Personal-Social	.6271	.5704
Full Scale	.8219	.8057

all p's < .001

Table 15

Parent vs. Teacher Developmental Age Mean Scores

Domains	Parent Mean DA	Teacher Mean DA	t Value	Significance Level
GD	38.14	35.79	3.30	$p < .001$
GM	38.37	37.41	0.67	$p > .10$
FM	45.39	45.20	0.17	$p > .10$
EL	29.88	27.78	3.83	$p < .001$
CC	38.36	35.01	3.90	$p < .001$
SC	37.33	32.66	4.59	$p < .001$
SH	45.13	42.12	2.25	$p < .05$
PS	34.05	30.04	3.55	$p < .001$
FS	39.57	37.02	4.27	$p < .001$

correlated with the children's chronological ages. Again, all correlations were Pearson product moment r's. Developmental age estimates from parent and teacher reports were compared with SB MA's with dependent t tests to assess whether the two different methods for arriving at age estimates resulted in different values.

All correlations between MCDI DA's from parent reports and SB MA's were highly significant (all p's < .001). The General Development DA and the SB MA were the most highly correlated scores with r = .7534. The correlation between Gross Motor DA and SB MA was the lowest correlation with r = .3676 (see Table 16).

The correlation between MCDI DA's from teacher reports and SB MA's were also highly significant (all p's < .001). Again, the highest correlation was between General Development DA and SB MA (r = .8106) while the lowest correlation was between Gross Motor DA and SB MA with r = .4737 (see Table 16).

Correlations between MCDI DA's and SB chronological age show that developmental progression as measured on the MCDI and chronological age are significantly related. The MCDI scales that showed the highest relationship (p's < .001) between chronological age and DA's from both parent and teacher reports were General Development, Fine Motor, Comprehension-Conceptual, Self Help, and Full Scale (see Table 17).

Comparisons of mean MCDI DA's with mean SB MA's show that, on the whole, MCDI DA's are lower than SB MA's (see Table 18). The mean overall DA estimate from the MCDI for both parent and teacher reports, obtained from the General Development scale, was significantly lower than the mean SB MA. The parent derived mean DA was approximately 4.5 months lower than the mean SB MA, and the teacher derived mean DA was approximately 7 months lower than the mean SB MA. The Self Help DA, from both parent and teacher derived reports was the only developmental scale which was not significantly different from the SB MA. The Fine Motor DA from both parent and teacher reports was the only scale score which was significantly higher than the SB MA. All other MCDI

Table 16

Correlations Between Parent and Teacher Derived

MCDI Developmental Ages and SB MA

MCDI Domains	Parent DAs with SB MA	Teacher DAs with SB MA
General Development	.7534	.8160
Gross Motor	.3676	.4737
Fine Motor	.6754	.7236
Expressive Language	.5668	.6830
Comprehension-Conceptual	.7406	.7662
Situation Comprehension	.5308	.5206
Self Help	.5949	.6271
Personal-Social	.5292	.5292
Full Scale	.7354	.7954

all p's < .001

Table 17

Correlations between Parent and Teacher Derived

MCDI Developmental Ages and SB CA

MCDI Domains	Parent DAs with SB CA	Teacher DAs with SB CA
General Development	.4736***	.4904***
Gross Motor	.1787*	.3205***
Fine Motor	.4483***	.4736***
Expressive Language	.2352**	.2331**
Comprehension-Conceptual	.4482***	.5164***
Situation Comprehension	.2280**	.3971***
Self Help	.4217***	.4245***
Personal-Social	.3017**	.2173**
Full Scale	.4183***	.4634***

***p < .001
**p < .02
*p < .05

Table 18

Mean Comparisons between MCDI DAs and SB MA

MCDI Domains	Parent Derived Mean DA	D[1]	MA	t-value	Significance Level
GD	38.14	-5	42.68	4.94	$p < .001$
GM	38.37	-5	42.68	2.64	$p < .01$
FM	45.39	+2	42.68	2.51	$p < .02$
EL	29.88	-13	42.68	13.58	$p < .001$
CC	38.36	-5	42.68	4.40	$p < .001$
SC	37.33	-6	42.68	4.68	$p < .001$
SH	45.13	+2	42.68	1.76	$p > .05$
PS	34.05	-9	42.68	7.38	$p < .001$
FS	39.57	-3	42.68	4.00	$p < .001$

MCDI Domains	Teacher Derived Mean DA	D[1]	MA	t-value	Significance Level
GD	35.79	-7	42.68	8.82	$p < .001$
GM	37.41	-6	42.68	3.54	$p < .001$
FM	45.20	+2	42.68	2.61	$p < .01$
EL	27.78	-15	42.68	18.20	$p < .001$
CC	35.01	-8	42.68	9.12	$p < .001$
SC	32.66	-10	42.68	9.32	$p < .001$
SH	42.12	-1	42.68	.46	$p > .10$
PS	30.04	-13	42.68	11.49	$p < .001$
FS	37.02	-6	42.68	8.34	$p < .001$

1. MCDI DA - SB MA

behavioral domain DA's were significantly different from and lower than the SB MA. The differences between the SB MA's and the MCDI DA's are largely due to cohort effects in MA/CA relationships in the 1972 revision. On the whole, SB DA's in this revision are about 6 months higher for this developmental age range.

Stanford Binet Subscales

Two experimental scales of the Stanford Binet were devised for this study--the Language Scale and the Nonlanguage Scale. The Language Scale consisted of all SB items which required a verbal response and the Nonlanguage Scale consisted of all items which did not require a language response. Mental ages were calculated for each scale by considering each scale as a shortened version of the entire SB. The correlation between the Language MA and the Nonlanguage MA was highly significant with r = .8189. Correlations between Language MA's and Nonlanguage MA's with MCDI DA's were also highly significant (see Table 19). General Development and Comprehension-Conceptual were the most highly correlated of the MCDI scales and the SB scales. The Expressive Language scale had the highest absolute difference in its correlations with the Language and Nonlanguage SB scales and it was more highly correlated with the Language Scale than the Nonlanguage Scale. All correlations were, again, Pearson product moment r's.

Mean comparisons were performed with dependent t tests to determine how Language and Nonlanguage MA's compared with each other, SB MA, and MCDI GD DA (see Table 20). The mean Nonlanguage MA was significantly higher than the mean Language MA, the Mean SB MA, and parent and teacher derived MCDI GD DA's. The mean Language MA was significantly lower than the mean Nonlanguage MA and the mean SB MA, but significantly higher than both parent and teacher derived MCDI GD DA means. Moderately retarded children do better on SB items which do not require a verbal response.

Down's Syndrome Diagnosis

The total sample of 93 children was divided into two independent groups.

Table 19

Correlations between Parent and Teacher Derived MCDI DAs with SB Language MA and SB Nonlanguage MA

MCDI Domains	Parent MCDI DAs with Language MA	Parent MCDI DAs with Nonlanguage MA	Teacher MCDI DAs with Language MA	Teacher MCDI DAs with Nonlanguage MA
GD	.7614	.7089	.8199	.7688
GM	.2876*	.3490	.4229	.4441
FM	.5773	.6850	.6636	.6912
EL	.6920	.4641	.7924	.5619
CC	.7520	.7010	.7826	.7274
SC	.5095	.4776	.4850	.4906
SH	.5432	.5875	.6052	.5982
PS	.4947	.5020	.5278	.4870
FS	.7282	.6912	.7983	.7362

All p's < .001 except "*", which was p < .003.

Table 20

Mean Comparisons with Language and Nonlanguage Scales

Comparisons	Means	t Value	Significance Level
Language MA with Nonlanguage MA	41.01 45,78	5.44	.001
Language MA with SB MA	41.01 42.68	2.47	.015
Nonlanguage MA with SB MA	45.78 42.68	8.13	.001
Parent MCDI GD DA with Language MA	38.14 41.01	2.82	.006
Teacher MCDI GD DA with Language MA	35.79 41.01	5.93	.001
Parent MCDI GD DA with Nonlanguage MA	38.14 45.78	7.43	.001
Teacher MCDI GD DA with Nonlanguage MA	35.79 45.78	11.12	.001

The Down's group consisted of 41 children. The non-Down's group was a heterogeneous group consisting of 52 moderately retarded children not diagnosed with Down's Syndrome. The average age of the Down's group was 96.6 months and the average age of the non-Down's group was 98 months, not significantly different. The mean scores for these two groups were compared with t tests for independent means on all parent derived MCDI domains, SB MA, and Language and Nonlanguage MA's.

Results of these comparisons revealed that the Down's group scored significantly higher than the non-Down's group ($p < .05$) on all MCDI scales except for the Expressive Language scale, where no difference was noted. No significant differences were found between these two groups on mean SB MA, mean Language MA, or mean Nonlanguage MA. The MCDI seemed to highlight the non-cognitive, non-language abilities of the children with Down's syndrome better than the Stanford Binet.

Discussion

Satisfactory interjudge (parent/teacher) reliability ($r > .80$) for the DA's based on the MCDI scales of General Development, Full Scale, and Comprehension-Conceptual was obtained. The rest of the scales had interjudge reliability coefficients in the .6 to .8 range, while only the Personal-Social scale was below .60. The DA's derived from parent reports on the MCDI scales of General Development, Expressive Language, Comprehension-Conceptual, Situation Comprehension, Self Help, Personal-Social, and Full Scale were significantly higher than the DA's derived from teacher reports. Differences ranged from 2.1 to 5.7 months. Only on the DA's derived from the Fine Motor and Gross Motor scales were the differences between parents and teachers not significant. Thus, parents and teachers ranked the children similarly, but parents saw their children performing more behvaviors than teachers. The MCDI is, of course, designed to be used by mothers and the norms are constructed from that source of data. Therefore,

when both reports are available, DA's should be calculated using mother-obtained data.

When the DA's calculated from mother and teacher MCDI responses are compared with MA's obtained from the Stanford Binet, teachers' results always have a higher correlation. However, since in general, all MCDI DA's were lower than SB MA's and since parent DA's were in general higher than teacher DA's, it is not surprising that DA's based on parents' reports come closer to SB MA's. Again, this would indicate that it is preferable to rely on mothers' reports; but if these are unobtainable, teachers' reports can be substituted with some caution.

With this sample of moderately retarded chidlren, as with the MCDI normative sample, the General Development scale was the best measure of DA in terms of reliability and validity. The GD scale had the highest interjudge reliability correlations and the highest correlation with Stanford Binet MA.

The older children in this moderately retarded sample had higher scores than the younger children, although the correlations of MA and DA with CA are smaller than those obtained with normal children, in which case the correlations would approach 1.0 with a perfectly reliable test. Again, the longest scales, the General Development scale (and the Full Scale) had the highest correlations. Expressive Language, Personal-Social, and Gross Motor DA's appeared to improve only slightly with age in this developmental age group. The Gross Motor DA's are approaching ceiling level in this developmental age group.

Although the Language and Nonlanguage SB scales were highly correlated with each other, there was some evidence with this sample that they were measuring differences in verbal vs. nonverbal tasks. The lowest correlation was between the Language score and the parent-reported Gross Motor score. The MCDI Expressive Language DA correlated .69 with the Language score, but only .46 with the Nonlanguage score, a difference replicated by the teachers' data. The Language scores were lower than the Nonlanguage scores for the sample as a whole. Thus,

this breakdown may be a helpful one for moderately retarded children.

The Down's Syndrome children did not differ from the non-Down's moderately retarded children on any of the Stanford Binet measures. Developmental ages from the MCDI, however, were consistently higher for the Down's Syndorme children on all MCDI scales, except for the Expressive Language scale. Therefore, although there is no difference between these two groups' performance on standard intellectual tasks, parents rated their Down's Syndrome children as performing a greater range of behaviors, except in expressive language skills, than their moderately retarded non-Down's Syndrome children.

References

Ireton, H. & Thwing, E. J. The Minnesota Child Development Inventory. Minneapolis, MI; Behavior Science Systems, Inc., 1974.

McCarthy, D. McCarthy Scales of Children's Abilities. New York: Psychological Corporation, 1970, 1972.

Terman, L. & Merrill, M. Stanford-Binet Intelligence Scale. Boston: Houghton Mifflin Company, 1972.

Wechsler, D. Wechsler Intelligence Scale for Children. New York: Psychological Corporation, 1949.